NPS55-86-023

NAVAL POSTGRADUATE SCHOOL

Monterey, California

TECHNICAL

STATISTICAL ASPECTS OF THE F/A-18
AGE EXPLORATION PROGRAM

GLENN F. LINDSAY

NOVEMBER 1986

Approved for public release; distribution unlimited.

Prepared for:
Naval Air Rework Facility
North Island, San Diego, CA 92135

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral R. C. Austin
Superintendent

D. A. Schrady
Provost

This report was prepared by:

UNCLASSIFIED

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S): NPS55-86-023	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION: Naval Postgraduate School; 6b OFFICE SYMBOL (If applicable): Code 55	7a NAME OF MONITORING ORGANIZATION
6c ADDRESS (City, State, and ZIP Code): Monterey, CA 93943-5000	7b ADDRESS (City, State, and ZIP Code)
8a NAME OF FUNDING/SPONSORING ORGANIZATION: Naval Air Rework Facility; 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N65888-86-WR00069
8c ADDRESS (City, State, and ZIP Code): North Island, San Diego, CA 92135	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO; PROJECT NO; TASK NO; WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
STATISTICAL ASPECTS OF THE F/A-18 AGE EXPLORATION PROGRAM

12 PERSONAL AUTHOR(S)
Lindsay, Glenn F.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical	FROM ____ TO ____	November 1986	50

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	sample number, inspection error, reliability estimate

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Selected statistical features of the AGE Exploration Program for F/A-18 aircraft are examined with emphasis upon sample number and the impact of inspection errors upon resulting reliability estimates. The identification of aircraft populations targeted by samples of fleet leader aircraft is also discussed.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL: Glenn F. Lindsay	22b TELEPHONE (Include Area Code): (408)646-2688; 22c OFFICE SYMBOL: Code 55Ls

STATISTICAL ASPECTS OF THE F/A-18

AGE EXPLORATION PROGRAM

A Technical Report

September, 1986

Glenn F. Lindsay
Naval Postgraduate School
Monterey, California 93943

ABSTRACT

Selected statistical features of the Age Exploration Program for F/A-18 aircraft are examined with emphasis upon sample number and the impact of inspection errors upon resulting reliability estimates. The identification of aircraft populations targeted by samples of fleet leader aircraft is also discussed.

SUMMARY

Implementation of the AGE Exploration Program (AEP) for F/A-18 aircraft by the Naval Air Systems Command involves sampling fleet leader aircraft emphasizing inspection of selected structural components. Sample size, and the interpretation of sample results, are the subject of this report.

When the objective of sampling is reliability estimation, one can, in addition to single point estimates, construct confidence bounds for fleet reliability. These reflect the quality of the estimate in terms of how big a sample was taken. In AEP inspection to date, the usual sampling result is that no discrepancies are found, hence point estimates of reliability are 1.0. The functional relations and graphs developed in this report permit one to, for the case of a discrepancy-free sample, place a lower bound on fleet reliability as a function of how many aircraft were inspected.

During inspection, some discrepancies may go undiscovered. When this happens, sampling results overstate reliability. In this paper a method is developed to adjust sample size or reliability estimates to account for the chance of inspection error, and curves are provided to simplify this adjustment.

Since aircraft sampled in the Age Exploration Program are fleet leaders in terms of usage, they are not particularly representative of the F/A-18 fleet that exists at that point in time. However, they should be representative of F/A-18 aircraft as those aircraft reach the same usage level that characterized the sample. Careful identification of this future population increases future utilization of the reliability estimates from current AEP data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

STATISTICAL ASPECTS OF THE F/A-18
AGE EXPLORATION PROGRAM

The Naval Air Systems Command has established the Age Exploration Program (AEP) for F/A-18 aircraft using Reliability-Centered Maintenance procedures in an effort to reduce maintenance costs by specifying only maintenance insuring flight integrity. Among other features of this program, fleet leader aircraft are sampled on a regular basis, with emphasis on inspection of selected structural components. It is the size of this sample and the statistical interpretation of the resulting data that form the subject of this report.

Since a stated purpose of sampling in AEP is the estimation of fleet reliability, this report first discusses reliability estimation, with emphasis on the relationship between sample size and the goodness of the estimate, when the measure of effectiveness for the estimate is confidence interval size. Curves are provided for determining the lower 95% bound on reliability when no discrepancies are found in the sample.

The next section of this report considers the effect of inspection error on reliability estimation. Concepts from signal detection theory are employed to develop

relationships which may be used so as to partially compensate for these errors. Curves are provided which permit adjustment of reliability confidence bounds when discrepancies may be undiscovered during inspection of the aircraft component.

The relationship of sample and population is examined. Aircraft inspected under AEP are fleet leaders as identified by several measures of wear and tear, and usage. Identification of a population from which these aircraft may be considered a representative sample is important, since it is to this population that the reliability estimates will apply. After suggesting how such a population might be defined, the report concludes with a brief review of previous studies addressing AEP sampling.

A. Reliability Estimation and Confidence Bounds

In sampling to estimate the proportion of a population's items that possess some stated attribute, the standard approach is to sample n items, count x possessing the attribute, and then use the sample proportion x/n as the estimate of the unknown population proportion. The n trials or observations are assumed to be independent of each other, and the chance of the attribute being present should be the same in each trial.

In addition to the point estimate x/n, one can also

construct a useful interval estimate which will place a lower bound on the unknown proportion. This lower bound is computed from the data in such a way that there will be a 95% chance that the bound will indeed be below the unknown proportion. The result, for example, might say that we are 95% certain that a component's reliability is greater than 0.88, where the lower bound 0.88 was computed from the data resulting from sampling. The confidence interval method has the virtue of reflecting the size of the sample, and thus the accuracy of the estimate.

Applying these ideas to reliability estimation is quite straightforward. We are concerned with an aircraft population of finite size, where the unknown reliability is the proportion of aircraft in the population that do not possess a discrepancy at a particular inspection site on the aircraft, such as the stabilator attach fitting.

If we sample (inspect) n aircraft and find x with discrepancies at the inspection site, then our point estimate for population reliability is

$$\hat{R} = \frac{n-x}{n} \quad . \tag{1}$$

Statistical work with this kind of estimate usually assumes that the sample was taken randomly from the population, and that sampling was without replacement or from an infinite population.[1]

In application, a difficulty with a point estimate such as (1) is that the estimate $\hat{R}$ itself does not provide any measure of its closeness to the true reliability R. Finding no discrepancies in a sample of ten items yields the same estimate of reliability as finding no discrepancies in a sample of 100 items. In both cases the reliability estimate is $\hat{R} = 1.0$, but clearly we have more confidence in the latter. Simply knowing that bigger samples give better estimates (in terms of accuracy) does not offer guidance regarding how big a sample one ought to take. To relate sample size to the goodness of the estimate requires a measure of the effectiveness of the estimate, and this may be found through the application of confidence intervals instead of point estimates.

The best-known procedure for developing confidence intervals for proportions is attributed to Clopper and Pearson, and we shall follow their approach.[2] We seek a 95% lower bounded confidence interval for reliability. This means that we wish to use the data from the sample to construct a lower bound for the unknown population reliability, and that this lower bound should be such that we are 95% certain that it is less than the population reliability R. Thus from the sample data, we wish to find a lower bound such that the probability that (Lower Bound $<$ R) is 0.95.

The value of Lower Bound is to be computed from the results of the sample, and we shall focus upon the AEP experiences to date where the sample contains no discrepancies. Thus $x = 0$, and $\hat{R} = 1.0$. From this sample result, the lower bound is determined by asking how low the population reliability could be while allowing a 5% chance of no discrepancies in the sample. This value of reliability will be the lower bound.

For reliability R and sample size n, the probability of no discrepancies in the sample is R^n. Accordingly, for a 5% chance of no discrepancies at our lower bound, we have from the binomial distribution

$$(\text{Lower Bound})^n = 1-0.95$$

or

$$\text{Lower Bound} = (1-0.95)^{1/n} \tag{2}$$

as our 95% lower confidence bound on reliability R when the sample result is no discrepancies. A similar derivation could be made when the result is one discrepancy in the sample, two discrepancies, and so on.

From (2) it is clear that with a discrepancy-free sample, our lower bound on population reliability R increases with sample size. This is illustrated numerically by the values in Table 1, showing lower bounds associated with various sample sizes.

TABLE 1. Sample Size and 95% Lower Confidence Bounds on Reliability When No Discrepancies are found in the Sample

Sample Size	Lower Bound on Reliability
10	0.741
15	0.819
20	0.861
25	0.887
30	0.905
100	0.970

In application, we could say that if we took a sample of size 25 and found no discrepancies, we would be 95% certain that population reliability was greater than 0.887. Stated differently, we would have 95% confidence that no more than 13.3% of fleet aircraft of this age will have the discrepancy. A plot showing lower bounds as a function of sample size for the no-discrepancy case is given in Figure 1.

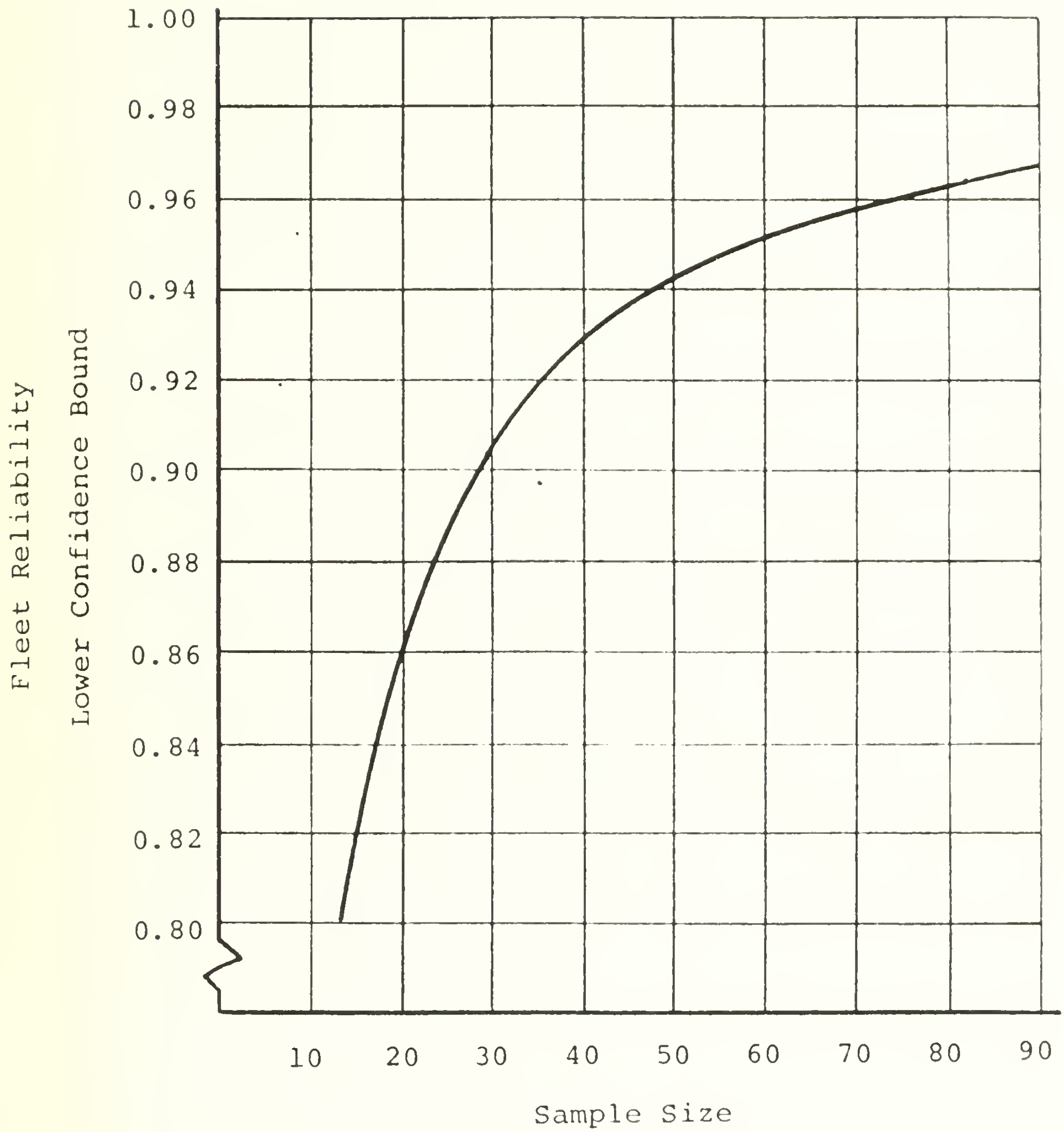

FIGURE 1. Lower 95% Confidence Bounds for Fleet Reliability when no Discrepencies are found in the Sample.

B. Effects of Inspection Errors on Reliability Estimation

The foregoing discussion of point estimates and lower confidence bounds for reliability tacitly assumed that each observation was correct, in the sense that the determination that an item did or did not possess a discrepancy was without error. The body of literature on inspection errors in non-destructive inspection is a growing one, and there seems to be increasing concern that the assumption of error-free performance on the part of inspectors, inspection hardware, and inspection procedures is questionable.[3,4,5,6] In this section we shall discuss the impact of errors on reliability estimates, and develop a way of adjusting the estimate to partially compensate for errors in data.

In a trial to determine whether an attribute is present, two kinds of errors are possible. The observation may be that the attribute is present when in fact it is not, or, the observation may be that the attribute is not present when in fact it is. Error performance on the part of the inspection process may be expressed for our reliability estimation case in the signal detection theory manner by two measures:[7]

p_d as the probability of a correct detection of a discrepancy, i.e., the inspection concludes that a discrepancy is present given there truly is a discrepancy, and

p_{fa} as the probability of a false alarm, i.e., the inspection concludes that a discrepancy is present when in fact there is none.

Using these two measures of detection performance, error-free inspection is described by

$$p_d = 1.0$$

and

$$p_{fa} = 0 \quad .$$

Suppose a population of N items contained A items with discrepancies and thus N-A good items, so that the population's true reliability would be

$$R = \frac{N - A}{N} \quad .$$

If we do 100% inspection (inspect every item in the population), we will on the average recognize a proportion p_d of the A items with discrepancies. Additionally, we will on the average declare a proportion p_{fa} of the good items to have discrepancies. In total, then, our average count of items with discrepancies would be

$$p_d A + p_{fa}(N-A) \quad .$$

From this, our statement of observed reliability after

100% inspection would be

$$R_{obs} = \frac{N - (p_d A + p_{fa}(N-A))}{N} \quad .$$

With some direct algebra, we have

$$R_{obs} = 1 - p_d(1-R) - p_{fa}R \quad ,$$

or

$$R_{obs} = 1 - p_d + R(p_d - p_{fa}) \quad . \tag{3}$$

Thus from (3) we see that the average value of observed reliability in 100% inspection is a linear function of the true reliability R. An example of the relative importance of the two kinds of inspection errors is shown in Table 2, for inspection error performance of the order of $p_d = 0.9$, and $p_{fa} = 0.1$.

TABLE 2. Examples of the Impact of Inspection Errors on Expected Observed Reliability in 100% Inspection.

	Expected Observed Reliability		
True Reliability	p_d=0.9 p_{fa}=0	p_d=1.0 p_{fa}=0.1	p_d=0.9 p_{fa}=0.1
1.00	1.000	0.900	0.900
0.95	0.955	0.855	0.860
0.90	0.910	0.810	0.820
0.85	0.865	0.765	0.780
0.80	0.820	0.720	0.740

Returning to the relationship (3), if we solve it for actual reliability R, we have

$$R = \frac{p_d - (1 - R_{obs})}{p_d - p_{fa}} \quad . \tag{4}$$

It is important at this time to again emphasize that R_{obs} is an <u>average</u> or expected value. When errors are possible ($p_d < 1.0$ or $p_{fa} > 0$), doing 100% inspection on the same population several times would probably yield a different reliability value each time. Equation (3) refers to the average result, and it is this average or expected value that is the argument in (4).

Returning to the effects of inspection errors on sample results, it is tempting to use the function (4) as a way of adjusting sample reliability results $\hat{R}$ to account for possible errors. If we sample n items from the population, count x with discrepancies, and compute reliability estimate $\hat{R} = (n-x)/n$, we might improve the estimate by adjusting it for inspection errors via

$$\hat{R}_{adj} = \frac{p_d - (1 - \hat{R})}{p_d - p_{fa}} \quad . \tag{5}$$

Note that this requires prior estimates of p_d and p_{fa} if one wishes to adjust the sample reliability estimate to account for possible inspection errors.

While a seemingly reasonable format to "improve" estimates, application of (5) can lead to values for adjusted reliability $\hat{R}_{adj}$ which are negative, or which are greater than 1.0. This is because we have replaced the mean or average value of observed reliability in (4) by our direct reliability estimate $\hat{R}$, which is a random variable. In small samples from the same population, $\hat{R}$ could be very large, or very small. We can generally say that our adjusted reliability estimate will be in the range

$$0 \leq R_{adj} \leq 1.0$$

when

$$(1 - p_d) \leq \hat{R} \leq (1 - p_{fa}) \quad .$$

A case of interest in the Age Exploration Program is that where p_{fa} is presumed to be small or negligible because discrepancies discovered by one inspection method are "confirmed" by a different inspection method. If we assume $p_{fa} = 0$, then with an estimate of discrepancy detection probability p_d, we would from (5) adjust our reliability estimate by

$$\hat{R}_{adj} = 1 - \frac{(1 - \hat{R})}{p_d} \quad . \tag{6}$$

Numerical examples for various p_d's are shown in Table 3, where we can see the magnitude of adjustment or correction of reliability estimates that would occur when we feel that discrepancy detection is imperfect.

TABLE 3. Reliability Point Estimates Adjusted for Discrepancy Detection Probabilities p_d, where $p_{fa} = 0$

Reliability Estimate from Sample	Adjusted Estimate $\hat{R}_{adj}$				
$\hat{R}$	$p_d=0.9$	$p_d=0.8$	$p_d=0.7$	$p_d=0.6$	$p_d=0.5$
0.5	0.44	0.37	0.29	0.17	0
0.6	0.55	0.50	0.43	0.33	0.20
0.7	0.66	0.62	0.57	0.50	0.40
0.8	0.77	0.75	0.71	0.67	0.60
0.9	0.89	0.87	0.86	0.83	0.80
1.0	1.00	1.00	1.00	1.00	1.00

The same adjustment can be made to our estimate of reliability using confidence intervals. Figure 2 shows the lower 95% confidence bounds on reliability adjusted for

various values of discrepancy detection probabilities p_d, for the case where no discrepancies were found in the sample Thus if we felt that the chance of finding a discrepancy in inspection was $p_d = 0.8$ and had found no discrepancies in a sample of size 30, we might state with 95% certainty that the population reliability was greater than 0.88. In other words, we have 95% confidence that no more than 12% of fleet aircraft at this age will have the discrepancy.

Using Figure 2 it is possible, of course, to make a reliability estimate before the entire sample of 30 is inspected. After the first ten aircraft were inspected our lower bound at $p_d = 0.8$ would be 0.68 for reliability. This estimate and the later one at n=30 are, of course, not independent.

Functionally, the curves in Figure 2 show

$$\text{Lower Bound}_{adj} = 1 - \frac{1 - (1 - 0.95)^{1/n}}{p_d} \quad . \tag{7}$$

Figures 3 and 4 provide the same information as Figure 2 for confidence bounds of 90%, and 99%, respectively.

95 Per Cent Bounds

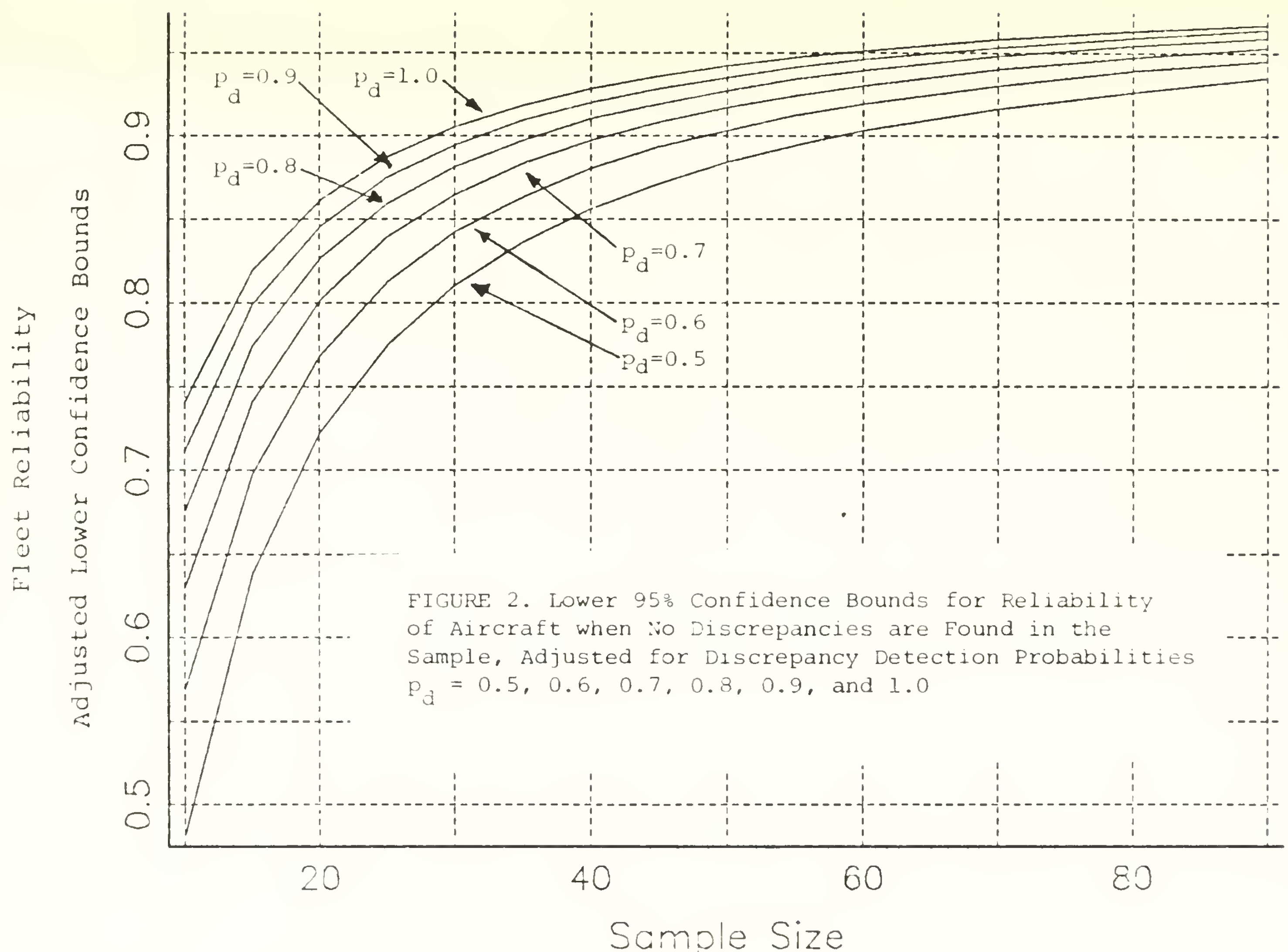

FIGURE 2. Lower 95% Confidence Bounds for Reliability of Aircraft when No Discrepancies are Found in the Sample, Adjusted for Discrepancy Detection Probabilities p_d = 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0

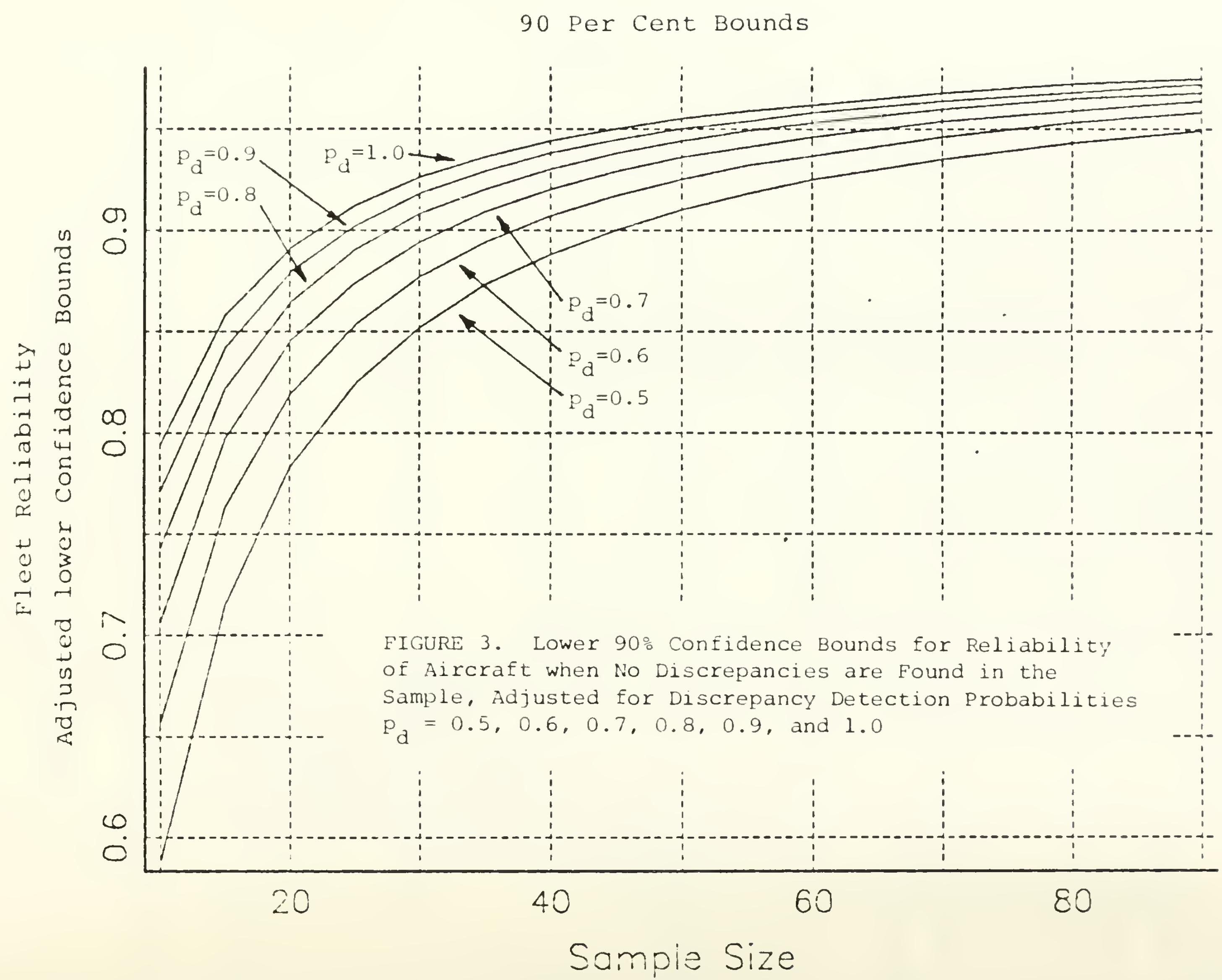

FIGURE 3. Lower 90% Confidence Bounds for Reliability of Aircraft when No Discrepancies are Found in the Sample, Adjusted for Discrepancy Detection Probabilities p_d = 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0

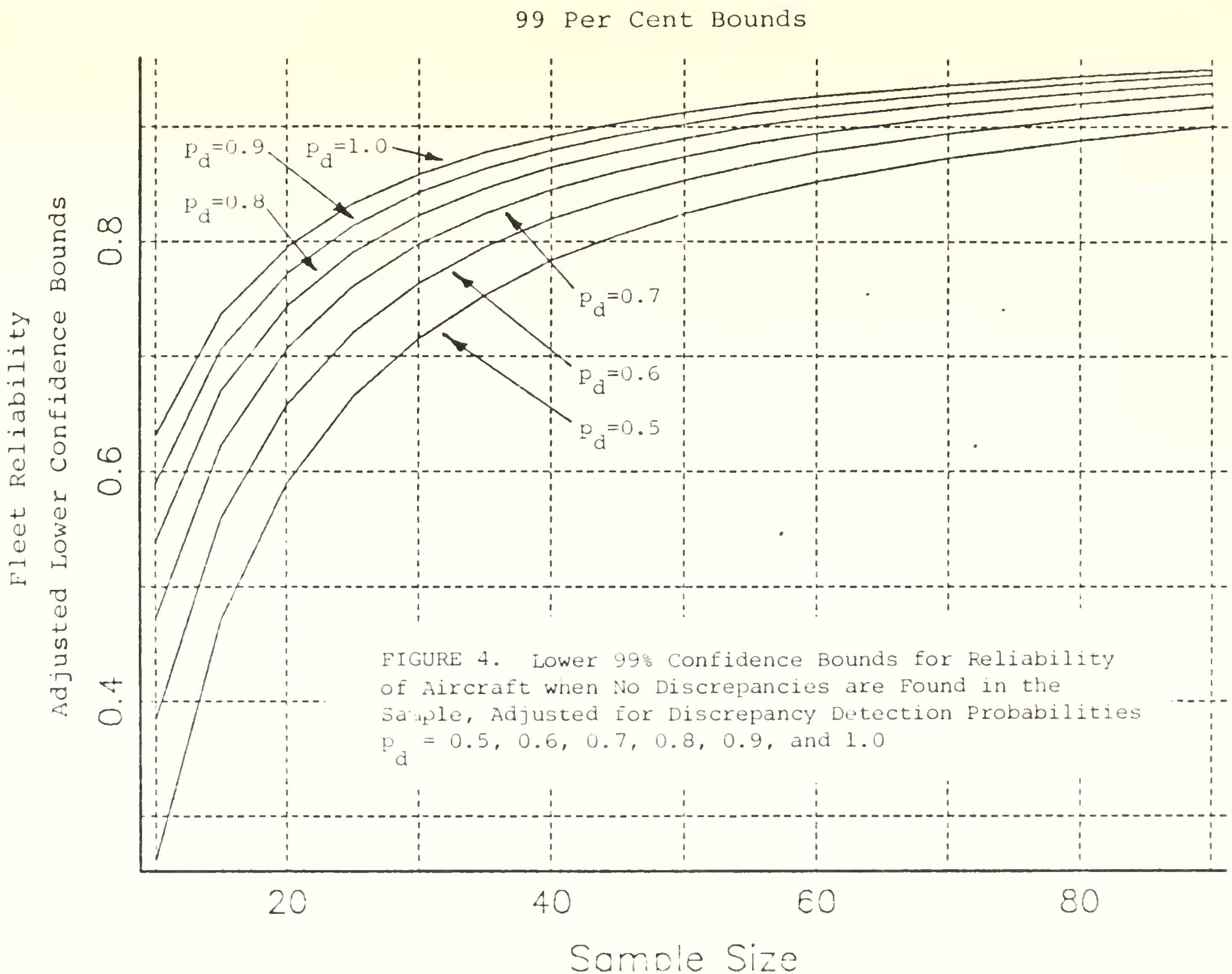

FIGURE 4. Lower 99% Confidence Bounds for Reliability of Aircraft when No Discrepancies are Found in the Sample, Adjusted for Discrepancy Detection Probabilities p_d = 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0

C. Accounting for Finite Populations

The foregoing work assumes that our samples come from populations of infinite size, or from sampling with replacement. This was inherent in our tacit use of the binomial probability distribution. In sampling in the Age Exploration Program, however, populations will be finite in size, and sampling is without replacement.

When populations are finite the correct probability distribution for the number x possessing the attribute out of a sample of size n is the hypergeometric distribution; this would have involved the use of population size in our calculations. It has been frequently demonstrated, however, that when the sample size is less than 10% of the population size, the hypergeometric is well approximated by the binomial distribution.[1]

Where the sample size exceeds 10% of the population, the lower bound value for reliability as computed earlier in this paper would understate the true value, and thus the error would be on the conservative side. For example, with a sample of 30 from a population of 300, the lower bound from the binomial is 0.9050, while the hypergeometric value for the lower bound is 0.9096. For aircraft populations of size 20, 30, 40, 50, and 100, sample size curves from the hypergeometric distribution are given in the Appendix to this report.

D. Characterizing the Sample

Because they consist of fleet leader aircraft, the samples taken and inspected in the Age Exploration Program are not representative of the entire fleet of F/A 18 aircraft that exists at the time the sample is taken. Accordingly, it is necessary to identify or characterize the population for which reliability is being estimated, and thus for which the sample should be representative.

Aircraft which are chosen to be in the sample are selected on the basis of age or usage, as defined by one or more measures. Two examples of these measures are cumulative arrestments, and the current value of the wing root fatigue index. The reliability estimated from the sample should be applicable to aircraft when they reach the age range represented in the sample. Such a population does not exist at a point in time, indeed, some of the aircraft addressed may not have been built yet.

The sample in AEP is not a random one. (A random sample is one taken in such a way that each element of the population has an equal chance of being in the sample.) For our purposes, however, we will assume that the aircraft inspected are a representative sample of F/A 18 aircraft in the age range characterizing the sample. The practice of using a sample of today's items to make statistical

inferences about future similar items is widely followed in agricultural, biological, medical, and even military, experimental work.

E. Defining the Population for which Reliability is Estimated

Suppose only one measure of aircraft age is used to describe the 1987 AEP sample, and for discussion purposes, suppose that measure is wing root fatigue index. The sample then can be characterized as having wing root fatigue index values between F_1 and F_2, and it seems reasonable that our reliability estimate would then be applicable to a population of aircraft which also have wing root fatigue index values between F_1 and F_2. At some time in their lives, most fleet aircraft may, as they age, be members of this population. It is when they are at that "age" that the reliability estimate will be applicable to them.

F. Other Studies Seeking Sample Size

This report has treated the purpose of AEP inspection as estimation of reliability, and the work has centered upon relating the quality of such estimates to the number of aircraft sampled. Using the goodness of the estimate as the measure of effectiveness, procedures were developed for determining sample size, and also for the

inclusion of inspection error in finding final sample size and reliability estimate.

In the past, other measures of effectiveness have been used to propose sampling procedures and sample sizes for aircraft maintenance. These are briefly described and contrasted below.

MCAIR. In their 1983 report from McDonnell Aircraft Company, Smith and Swanson proposed an initial sample of size 22 for AEP.[8] This satisfied their criterion that if 10% of aircraft have discrepancies, there should be a chance of 0.9 that the sample will include one or more aircraft with discrepancies. Use of values other than 10% and 0.9 would have yielded different sample sizes. Their criterion assumes that a representative sample has come from an aircraft population having 10% with discrepancies. Since those in the sample are to be the most severely used aircraft, it is clear that the sample is not representative of the group of 450 aircraft to which it was restricted, but of a population of aircraft with similar usage. Applied to reliability estimation (assuming p_d=0.7), a sample of size 22 with no discrepancies found would give us 95% certainty that the reliability was greater than 0.82, in a population of similar age and use. After this initial sample, they suggest a sample from each of the two remaining sets of 450 aircraft employing

a procedure called Bayesian. This approach involves the assumption of a specific probability distribution for fleet reliability, prior to the actual sampling. This *a priori* distribution is then combined with the actual data from the sample to produce an *a posteriori* probability distribution of reliability. Their report does not indicate which *a priori* distribution they use, how it is to be combined with actual data, or properties of the results.

USAF. A different inspection criteria is used by the United States Air Force in their sample-based Analytical Condition Inspection (ACI) Program for the F-15 aircraft.[9] This procedure operates like statistical hypothesis tests applied as acceptance sampling or control charts. A double sampling procedure is used.[10] A sample of size 11 is taken. No action follows if no discrepancies are found. If exactly one discrepancy is found a second sample of size 13 is taken, and should it contain any discrepancies, corrective action follows. Corrective action also ensues if more than one discrepancy was found in the first sample. The action, no action, results of this sampling procedure place it in the realm of statistical hypothesis testing rather than estimation. For this program an operating characteristic curve could

be constructed showing the probabilities of no corrective action as a function of fleet reliability.[1] Using this data to estimate reliability leads to problems because of unequal sample sizes, making year to year results not comparable as point estimates if a second sample is periodically taken. When no discrepancies are found, the sample is of size 11 and we would on the basis of this be 95% certain that reliability is greater than 0.66; this assumes 70% detection probability in inspection. Sample data will, of course, accumulate from year to year.

NARF, North Island. In the 1982 report 001-82 for the NARF, North Island, J.D. Hayes employs "the level of confidence that the sample is analogous to a population which in fact has at least the specified reliability".[11] This statement, which has been discussed by Haff[12], appears to be a requirement statement by which a sample size can be deduced. Although the measure of sampling effectiveness is different, the equations which accompany the procedure produce sample size curves which, with a different interpretation, yield values similar to those in this report when $p_d=1.0$.

These three earlier studies may by summarized. MCAIR produced a sample size of 22 to satisfy a stated probability statement. The Air Force used a method mirroring statistical hypothesis testing for their

sampling procedure, which is directed toward corrective action rather than estimating reliability. The 1982 NARF report employed probability statements to produce expressions similar to those developed early in this report. None of the three studies explicitly considered the effects of inspection error on the data or on the needed sample size.

G. Concluding Remarks

Deciding on sample size for any empirical activity requires criteria or effectiveness measures by which the effects of various alternative sample sizes can be compared and judged. In this study we have taken the purpose of sampling to be that of generating estimates of reliability, and then used the goodness of the estimate (as measured by confidence interval size) as the criteria.

This permits the user through the figures and tables given in this report to evaluate and compare different sample numbers. If one wishes to determine a single number as sample size, an acceptable lower bound for the reliability estimate must also be given. If we say that with no discrepancies in the sample, we want to be 95% certain that fleet reliability is greater than X, then the required sample size value can readily be obtained from the given curves.

We have provided for the adjustment of the above values to account for possible inspection errors. Here, Figure 2 on Page 15 is probably most useful. The chances of errors are described by the probability of detecting an existing discrepancy. Often, in application, error possibilities are not taken into account because it is felt too difficult to estimate the detection probability. In this regard it should be pointed out that not taking error into account is equivalent to estimating $p_d = 1.0$, and if one feels errors are made, one should be able to formulate a better estimate of p_d.

From an estimation point of view, a crucial part of AEP sampling is identifying the population for which the samples are representative. It is hoped that the work presented in this report will assist in identifying that population, and will be useful to those who must interpret and apply the results of AEP sampling.

APPENDIX: SAMPLE SIZE FOR FINITE POPULATIONS

When the population is small so that the sample exceeds 10% of the population, the binomial distribution should no longer be used as an approximation to the hypergeometric distribution.[1] In this appendix we shall use the hypergeometric distribution to provide fleet reliability confidence bounds as a function of sample size for populations of size 20, 30, 40, 50, and 100 aircraft.

The hypergeometric probability distribution is

$$\mathrm{Prob}(x|n,m,N) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}, \tag{8}$$

where:

N is the number in the population,

m is the number in the population that possess the attribute,

n is the sample size, and

x is the number in the sample that possess the attribute.

Here, reliability is $R = m/N$.

Our case of interest is when no discrepancies are found in the sample. Here, $x = n$, and the probability of this from (8) is

$$\mathrm{Prob}(x=n \mid n,m,N) = \frac{m!(N-m)!}{(m-n)!N!} \quad . \qquad (9)$$

For a 95% lower confidence bound, this probability should equal 0.05 where the bound is m/N. However, we cannot find exact 95% lower confidence bounds solving

$$\mathrm{Prob}(x=n \mid n,m,N) = 0.05$$

for bound = m/N, since both m and N are integer valued. In a population of size $N = 20$, for example, $m = 0,1,2, \ldots, 19, 20$. Thus the number of possible reliability values for the population is finite, namely $N+1 = 21$ values.

Partial numerical results from searching for 90% and 95% lower confidence bounds for fleet reliability when fleet size is $N = 20$, are shown in Table 4. The values in the table are confidence levels for various lower reliability bounds and sample sizes. For example, with a sample of size 13 from a population of 20 aircraft, we have

$$\mathrm{Prob}(0.9 < \mathrm{Reliability}) = 0.889,$$

and

$$\mathrm{Prob}(0.85 < \mathrm{Reliability}) = 0.969 \quad .$$

TABLE 4. Examples of Probabilities Computed from the Hypergeometric Distribution when x=n and Population Size is N = 20.

Sample Size	m: 15 R: 0.75	16 0.80	17 0.85	18 0.90	19 0.95
6	.871				
7	.917				
8	.949	.898			
9	.970	.932			
10	.984	.957	.895		
11	.992	.974	.926		
12		.986	.951		
13		.993	.969	.889	
14			.982	.921	
15			.991	.947	
16				.968	
17				.984	
18				.995	.900
19					.950

Thus, exact 95% confidence bounds cannot in most cases be obtained.

Figure 5 shows approximate 95% lower confidence bounds for fleet reliability as a function on sample size, for populations of size 20, 30, 40, 50, and 100 aircraft. It can be seen that as population size grows, the number of possible reliability values grows, and the curves approach that of Figure 1 in the body of this report, where the binomial distribution was used. It should be pointed out again that because reliability has become a discrete parameter with a finite number of values, the plotted points rather than the curves are defined. Also, visible irregularities are present since exact 95% confidence levels could not be obtained.

Plotted points in Figures 6 through 10 adjust the fleet reliability bounds from Figure 5 to reflect the possibilities of undetected discrepancies. Figures 11 through 15 repeat Figures 6 through 10, but for 90 % confidence bounds rather than 95%.

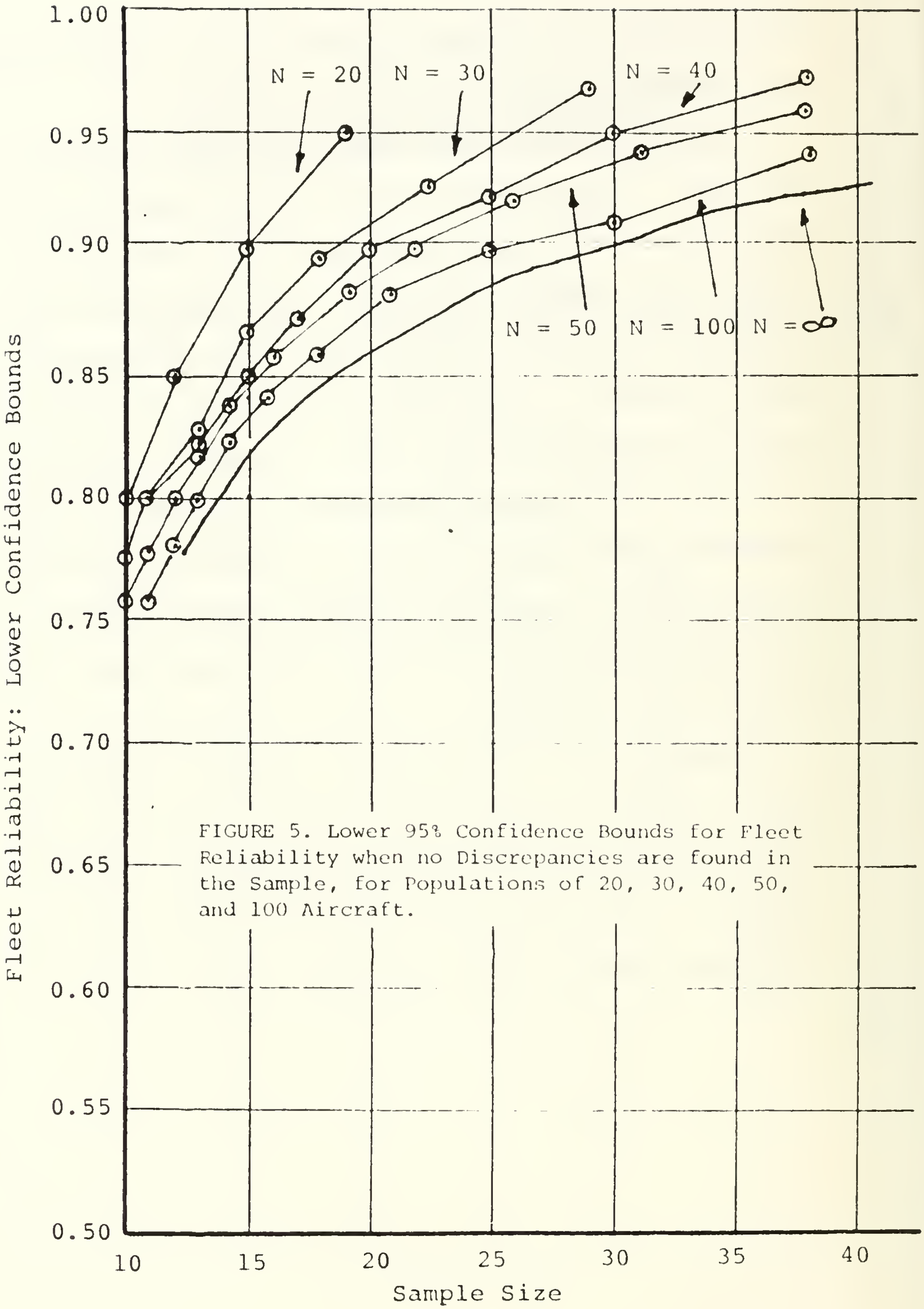

FIGURE 5. Lower 95% Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample, for Populations of 20, 30, 40, 50, and 100 Aircraft.

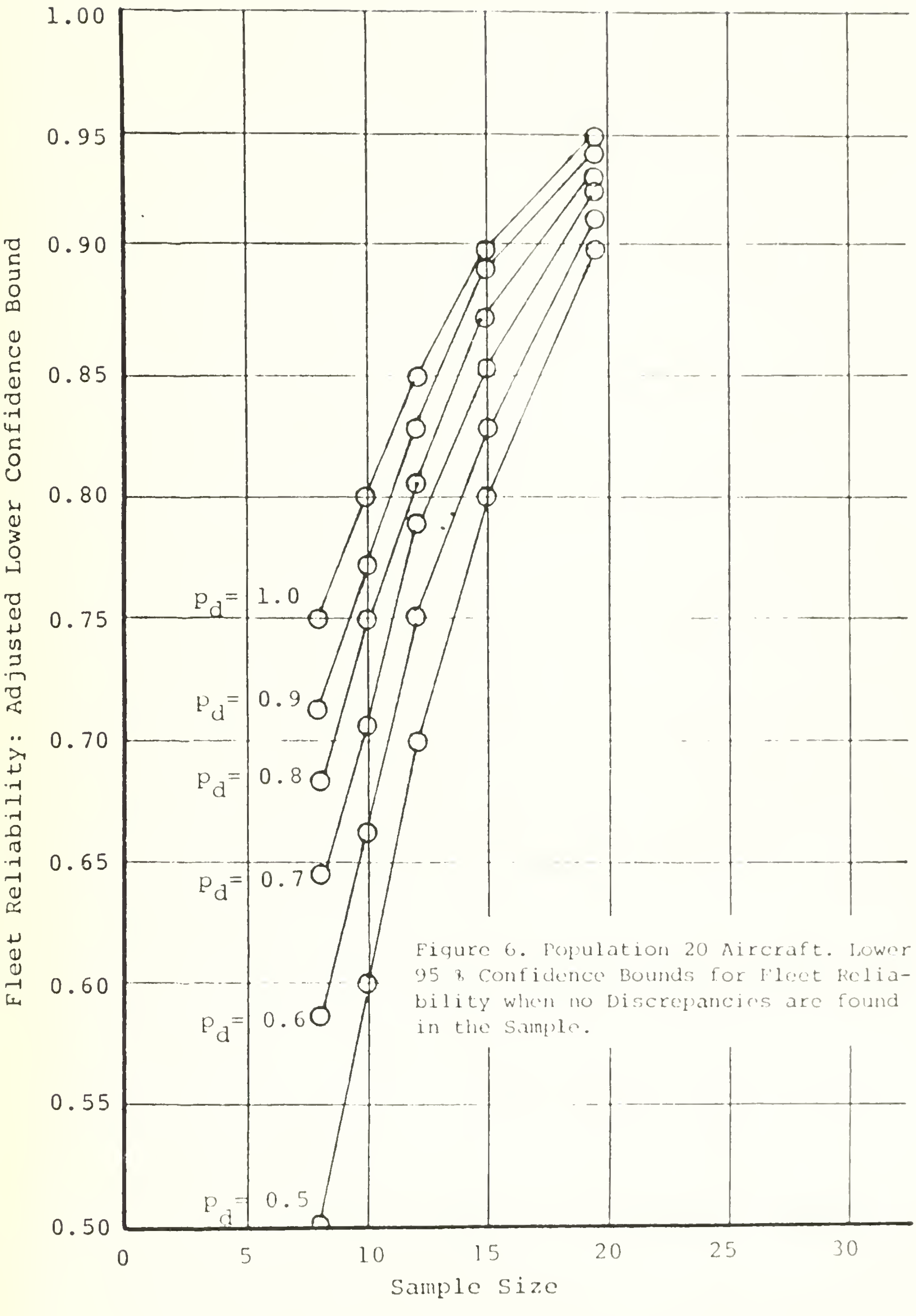

Figure 6. Population 20 Aircraft. Lower 95 % Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample.

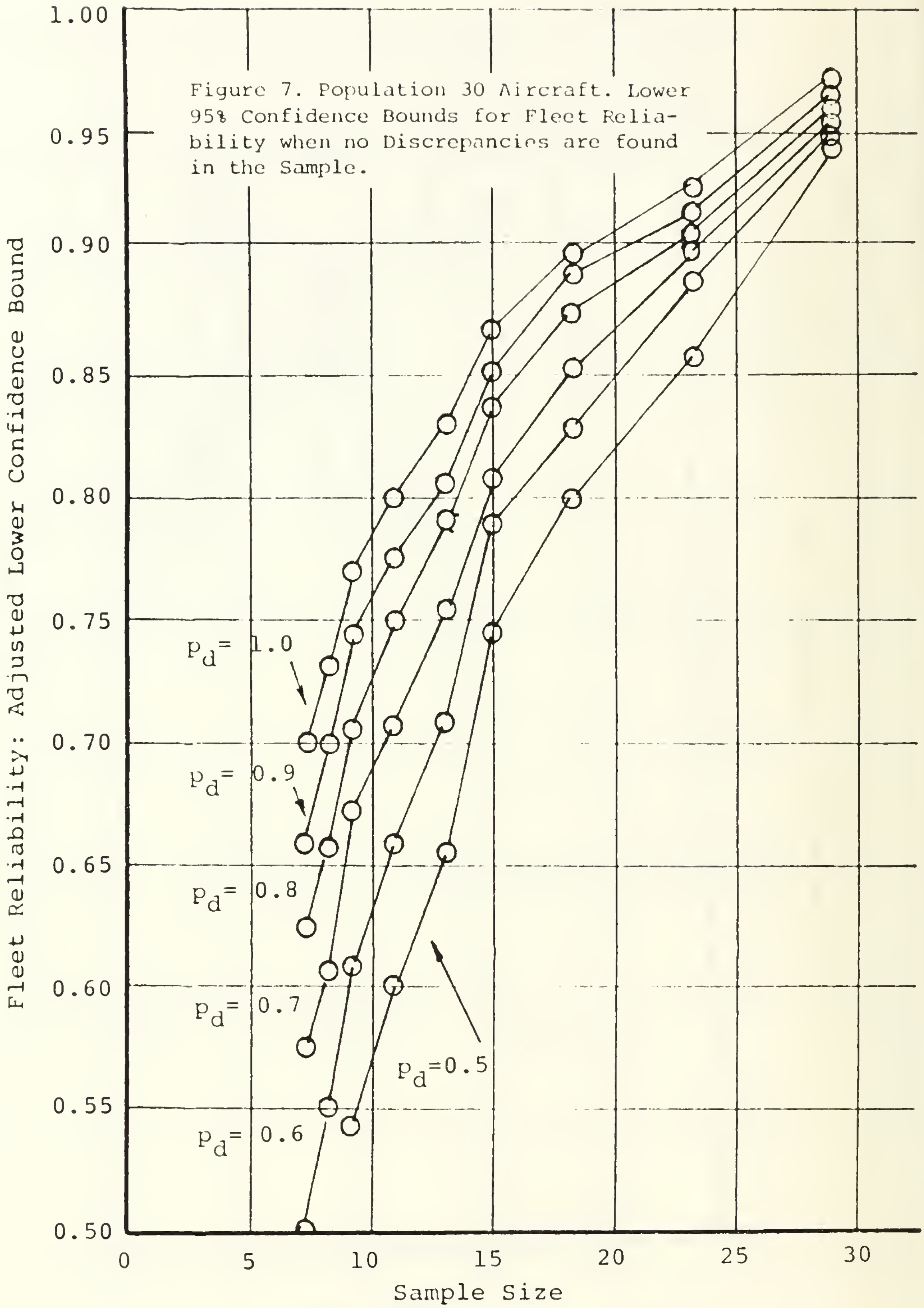

Figure 7. Population 30 Aircraft. Lower 95% Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample.

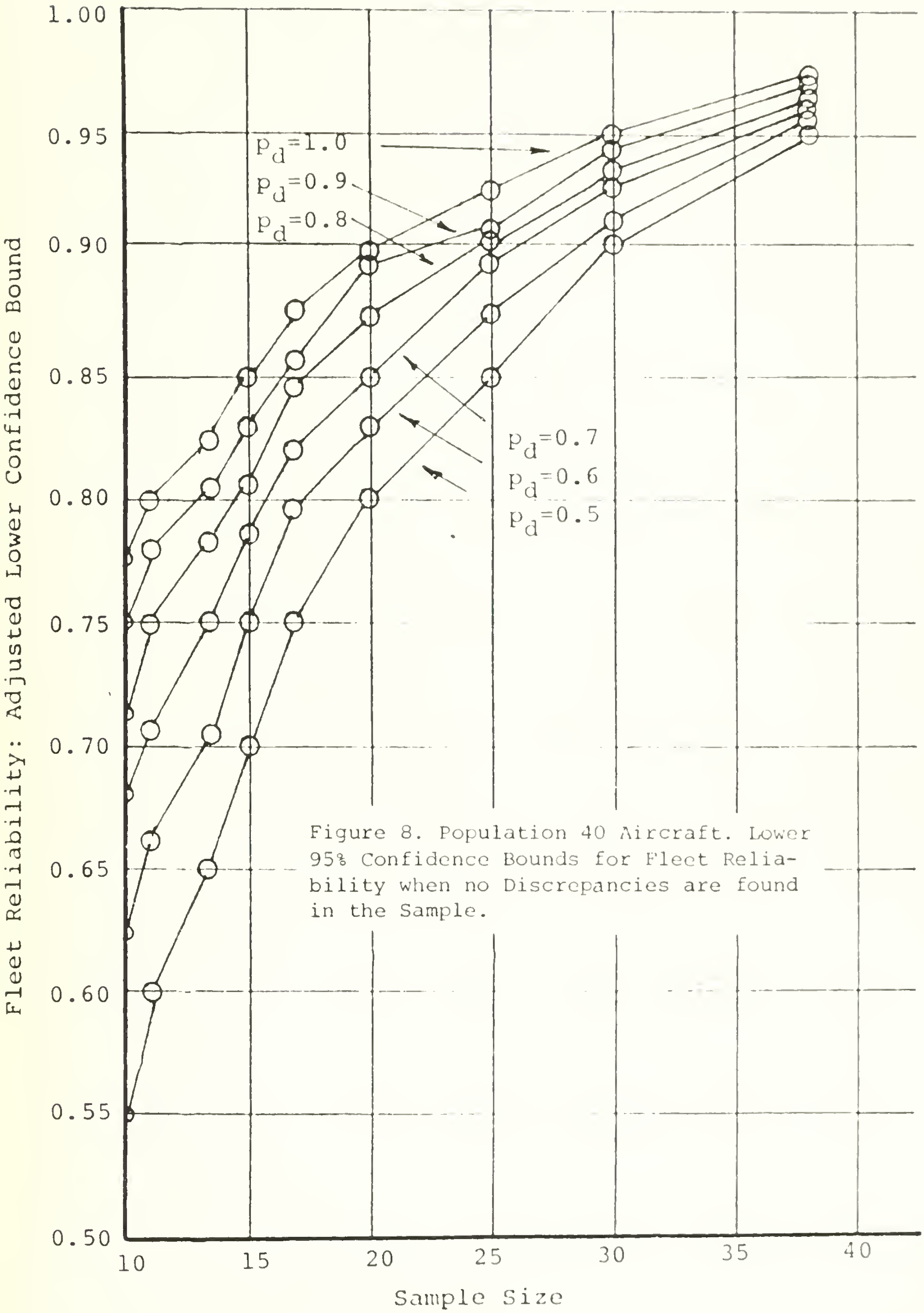

Figure 8. Population 40 Aircraft. Lower 95% Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample.

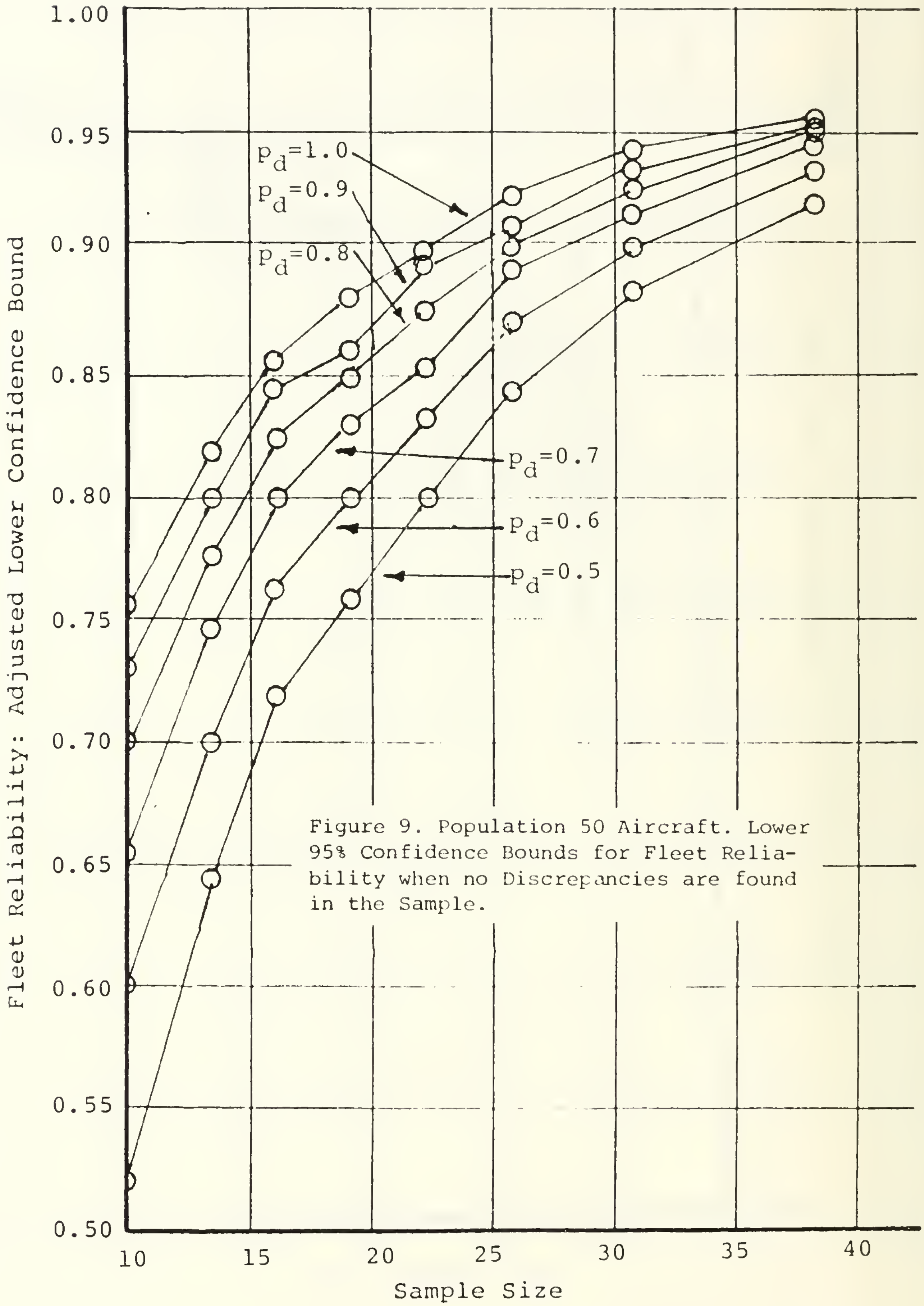

Figure 9. Population 50 Aircraft. Lower 95% Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample.

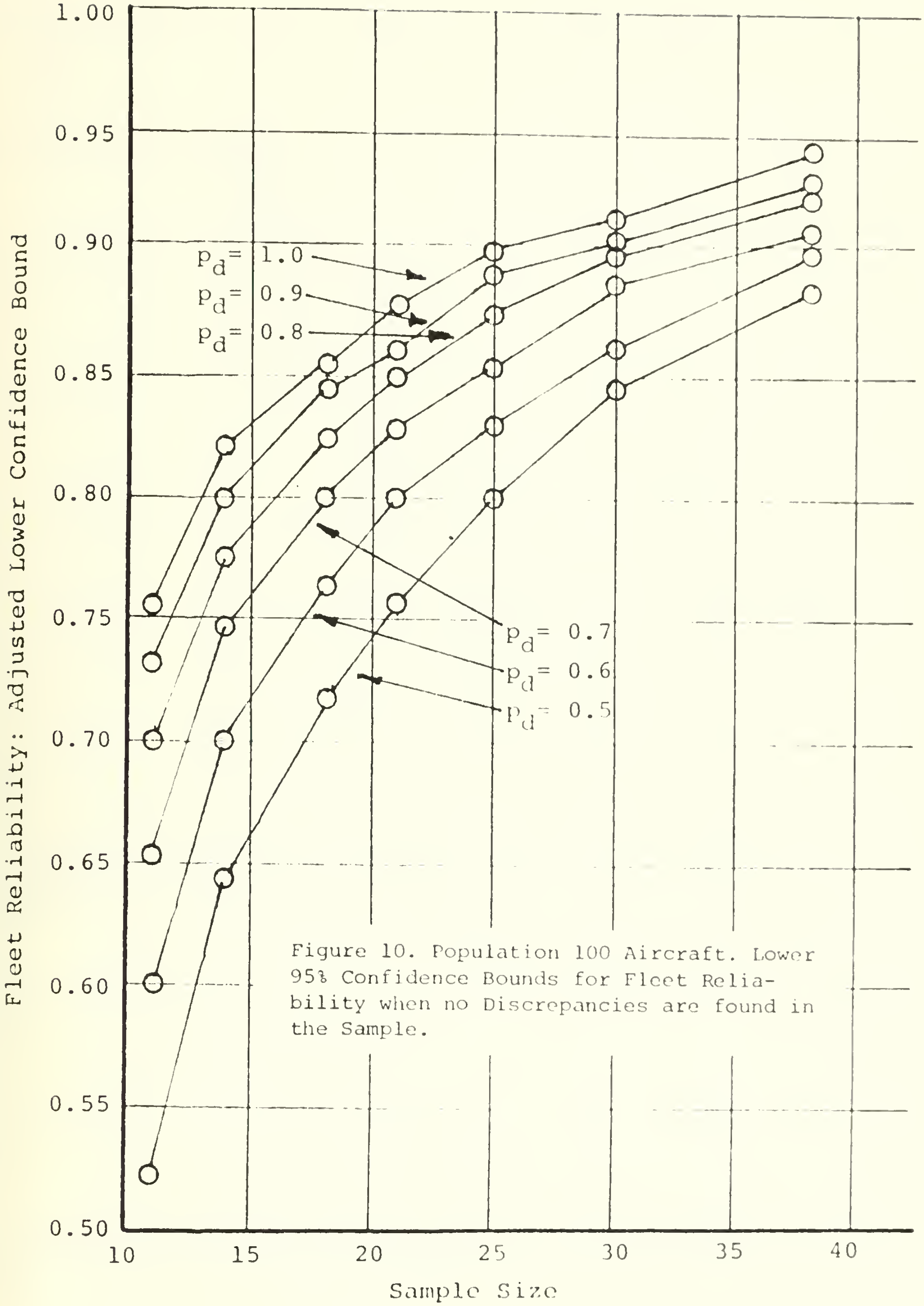

Figure 10. Population 100 Aircraft. Lower 95% Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample.

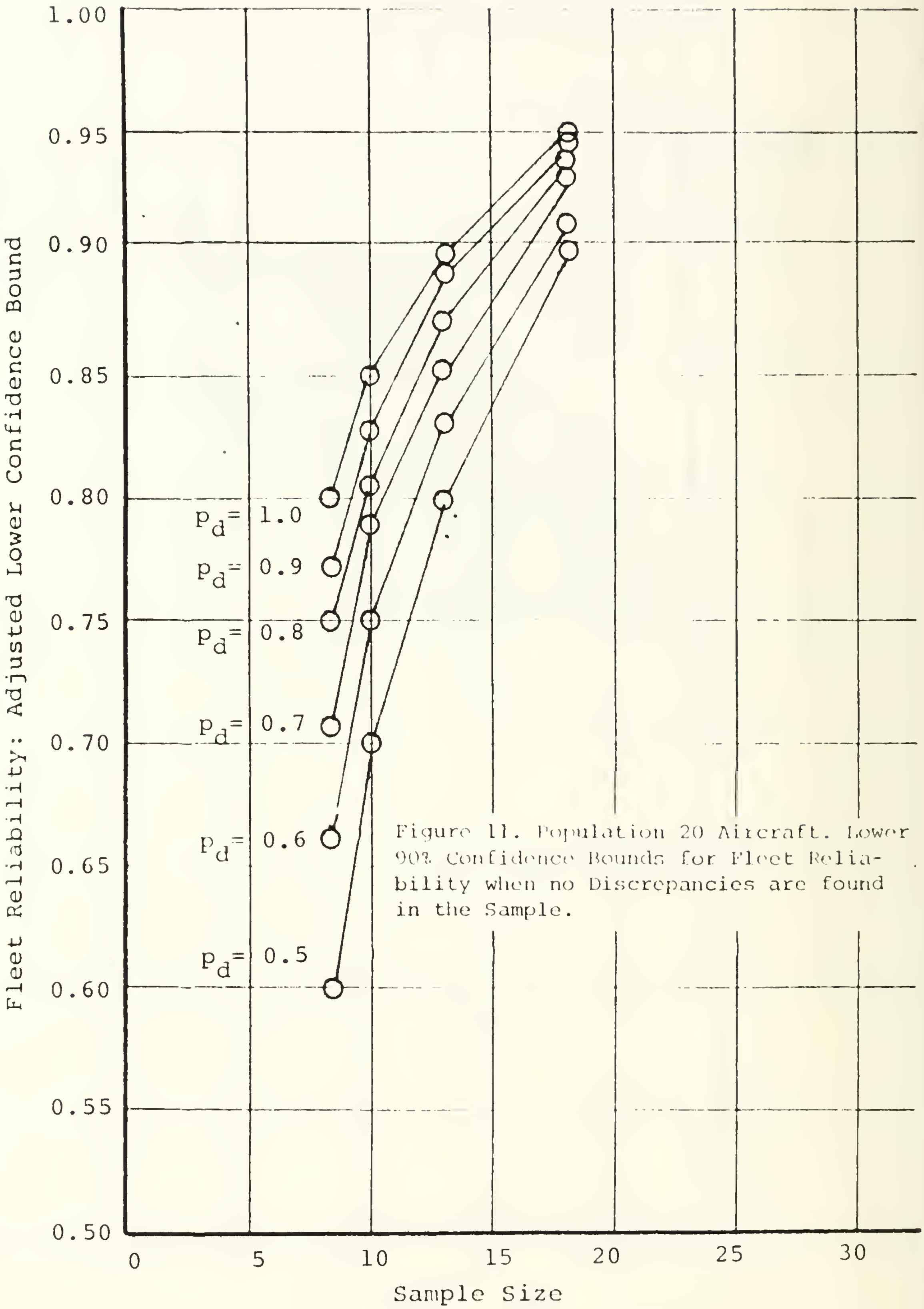

Figure 11. Population 20 Aircraft. Lower 90% Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample.

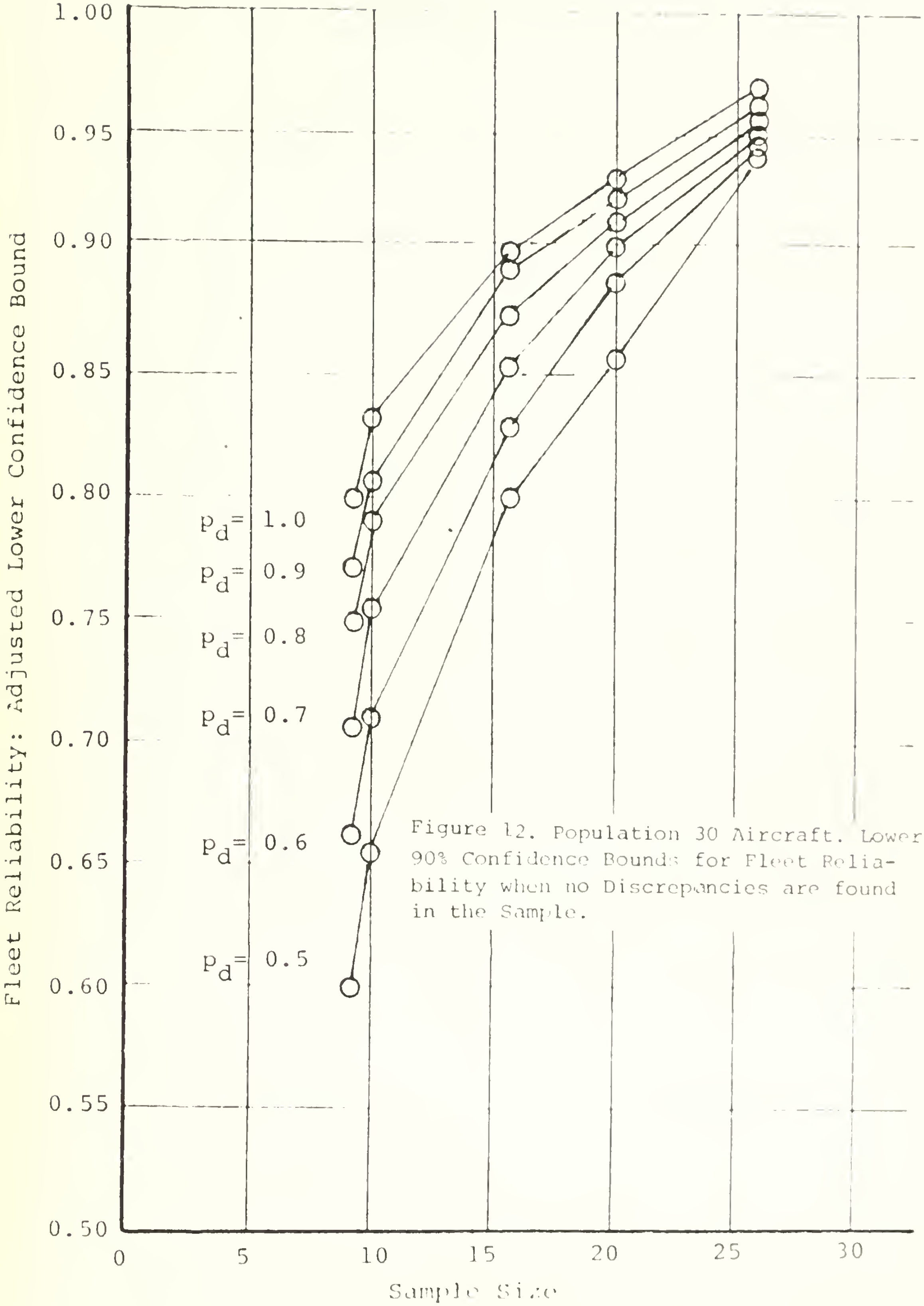

Figure 12. Population 30 Aircraft. Lower 90% Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample.

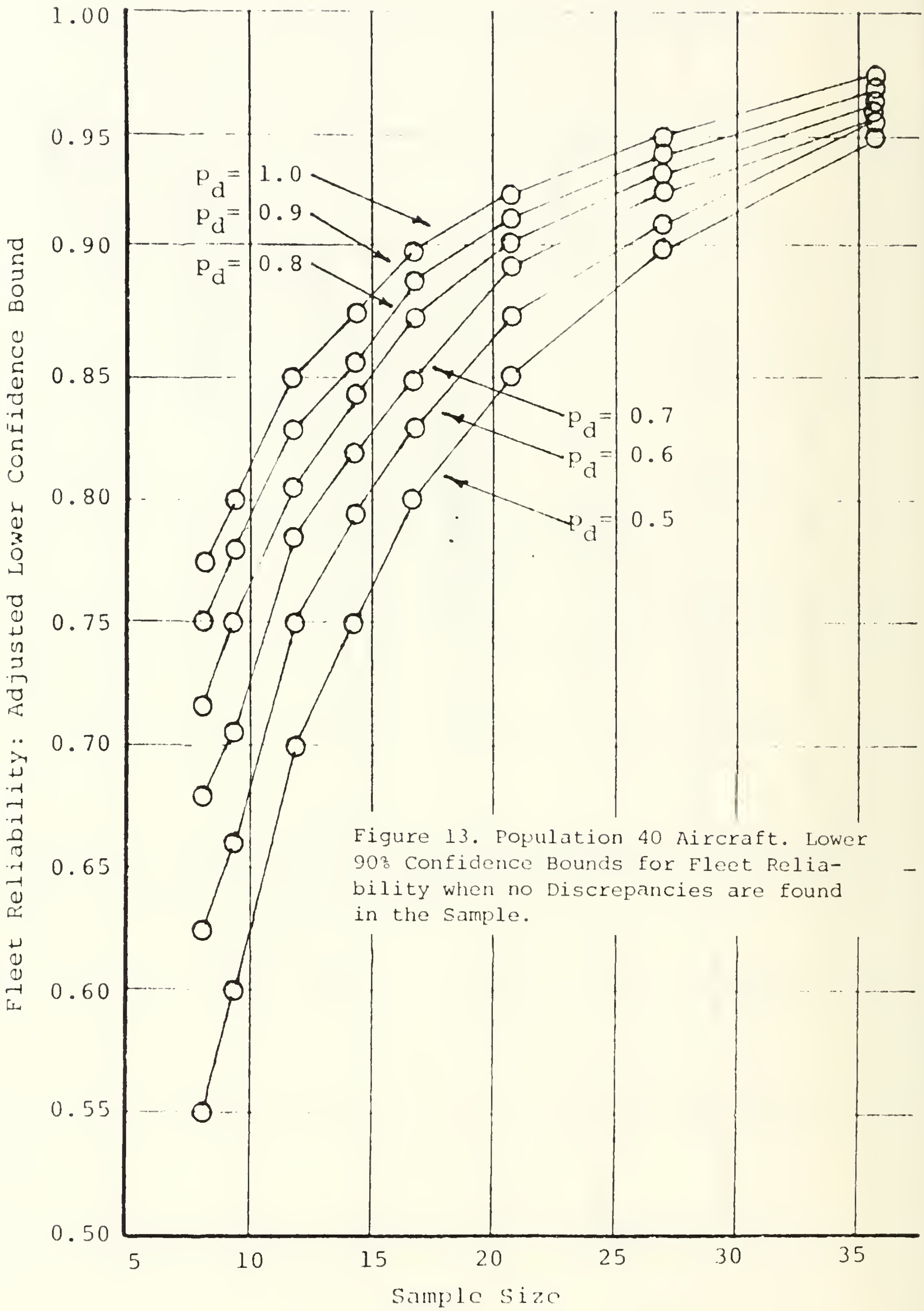

Figure 13. Population 40 Aircraft. Lower 90% Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample.

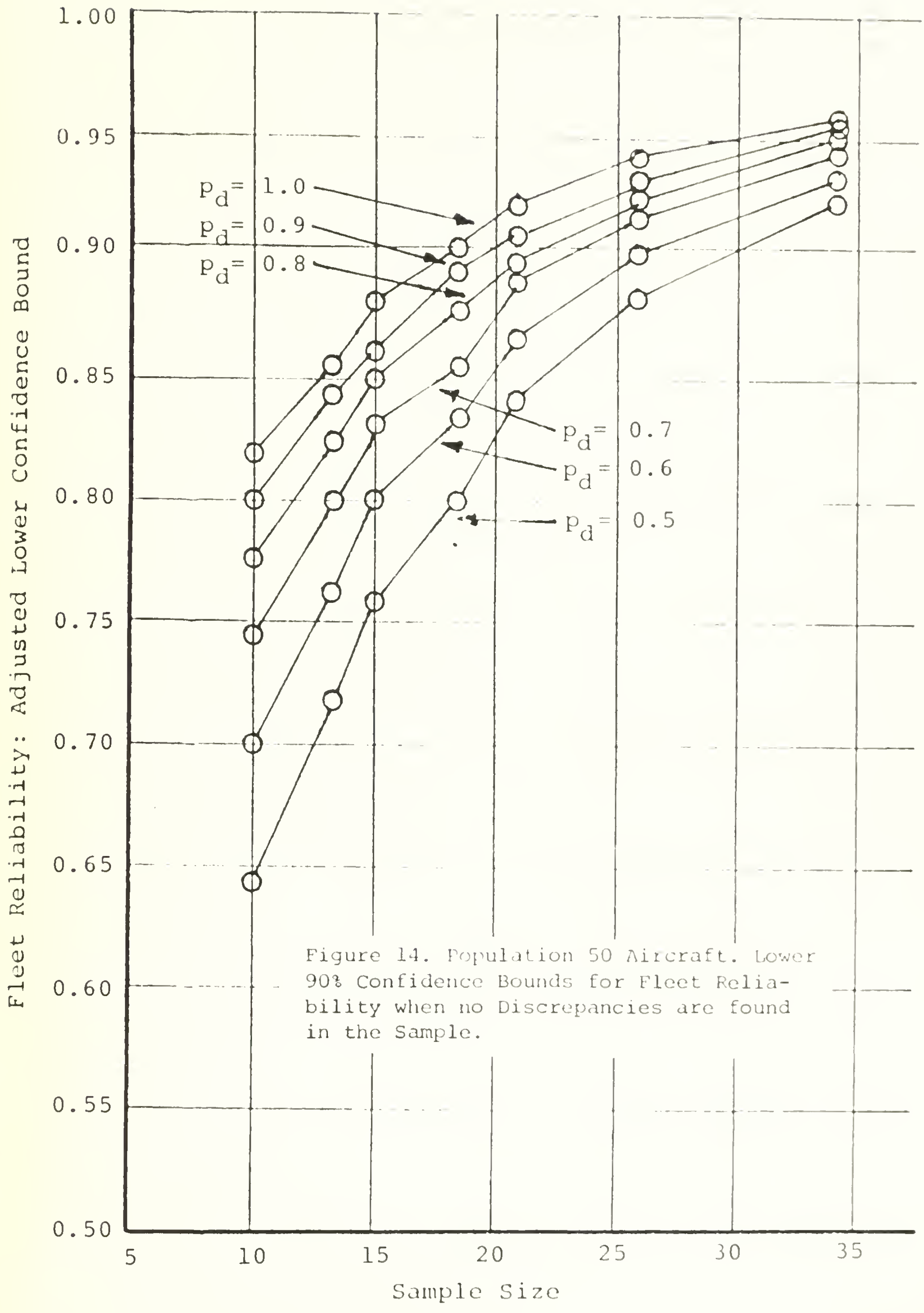

Figure 14. Population 50 Aircraft. Lower 90% Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample.

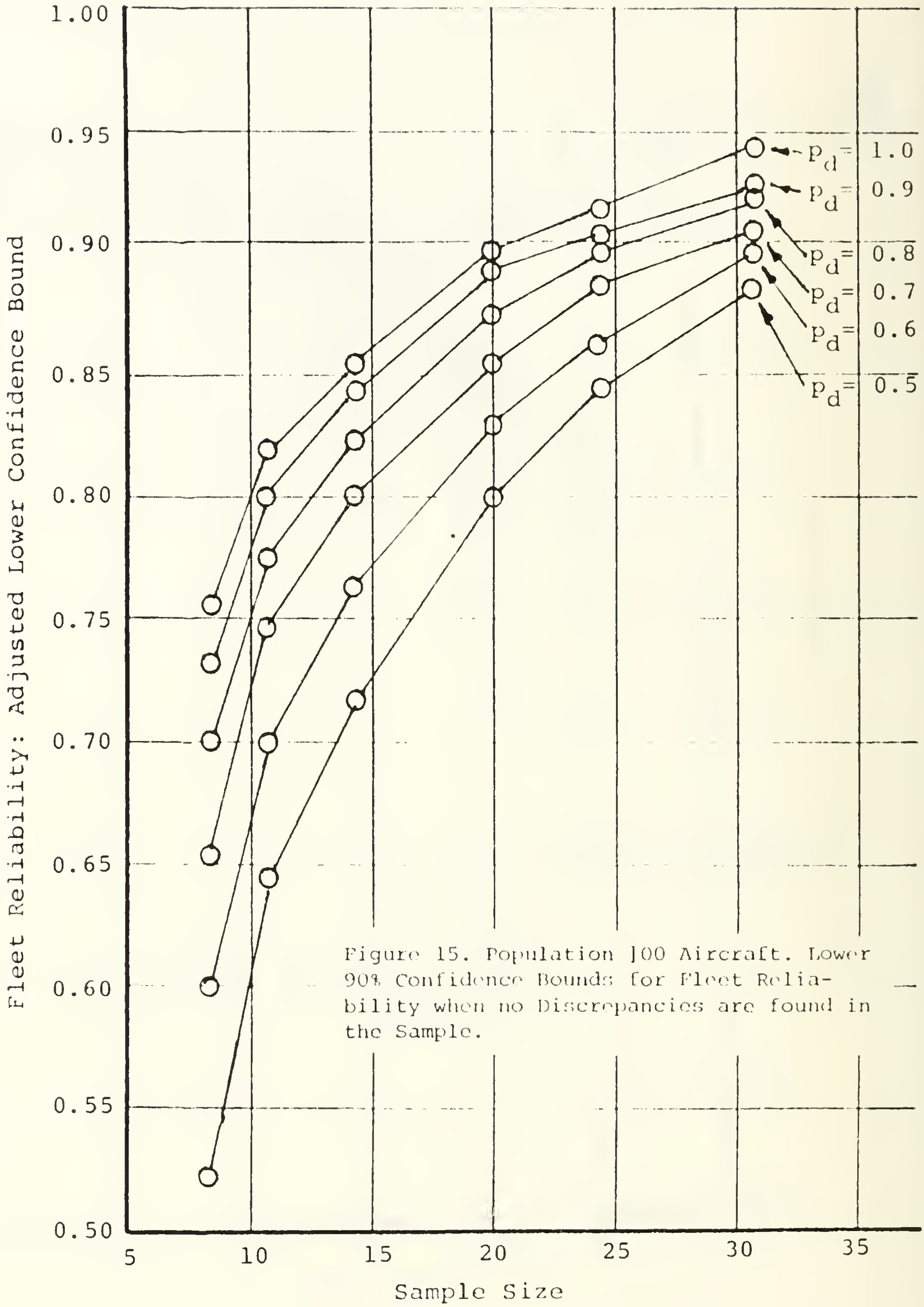

Figure 15. Population 100 Aircraft. Lower 90% Confidence Bounds for Fleet Reliability when no Discrepancies are found in the Sample.

LIST OF REFERENCES

1. Duncan, A.T., Quality Control and Industrial Statistics, 5th Edition, Irwin, 1986.

2. Clopper, C., and E. Pearson. "The Use of Confidence or Fiducial Limits Illustrated for the Case of the Binomial", Biometrika Vol. 26, pp404-413.

3. Drury, C.G., and J.G. Fox, Eds. Human Reliability in Quality Control. Halsted, 1975.

4. Jayachandran, T., and H.J. Larson. Statistical Analysis for the Nondestructive Inspection (NDI) Technical Proficiency Program. NPS Technical Report 53-83-0008, Naval Postgraduate School, June, 1983.

5. Lewis, W.H., and others. Reliability of Nondestructive Inspections. Prepared by the Lockheed Georgia Company, December, 1978.

6. Spanner, J.C., and others. Human Reliability Impact on Inservice Inspection, Vols. 1,2. Battelle Human Affairs Research Centers, Pacific Northwest Laboratory, March, 1986. (Prepared for the U.S. Nuclear Regulatory Commission).

7. Swets, J.A., W.P. Tanner, and T.G. Birdsall. "Decision Processes in Perception". Psychol. Rev. Vol 68. 301-310.

8. Smith, G.K., and H.L. Swanson. F/A 18A Age Exploration Program Depot Plan, Report P.S.1243, McDonnell Aircraft Company, St Louis, May, 1983.

9. Air Force Logistics Command, AFLC Regulations 66-28, Analytical Condition Inspection Program, 30 June, 1980.

10. Rose, Charles V. The Development of Scheduled Maintenance Programs for Naval Aircraft, M.Sc. Thesis, NPS, June, 1984.

REFERENCES, (CONTINUED)

11. Hayes, J.D. F/A 18 Age Exploration Program. Aero Engr Report 001-82, Naval Air Rework Facility, North Island, San Diego, January, 1982.

12. Haff, Leonard R. Preliminary Statistical Design and Analysis for the F/A 18 Age Exploration Program. Technical Report, Validity Corporation, Sept., 1981

DISTRIBUTION LIST

	NO. OF COPIES
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Library Code 0142 Naval Postgraduate School Monterey, CA 93943	2
Research Administration Code 012 Naval Postgraduate School Monterey, CA 93943	1
Library Code 55 Naval Postgraduate School Monterey, CA 93943	1
Professor Glenn F. Lindsay Department of Operations Research Code 55Ls Naval Postgraduate School Monterey, CA 93943	3
Naval Air Rework Facility, North Island NAVAIR Engineering Support Office Code 313, Building 6 San Diego, CA 92135	8